

UNIVERSITY OF
ILLINOIS LIBRARY
AT URBANA-CHAMPAIGN
STACKS

Digitized by the Internet Archive
in 2011 with funding from
University of Illinois Urbana-Champaign

<http://www.archive.org/details/critiqueofcartel599ulen>

Faculty Working Papers

A CRITIQUE OF CARTEL THEORY

Thomas S. Ulen, Assistant Professor,
Department of Economics

#599

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

August 16, 1979

A CRITIQUE OF CARTEL THEORY

Thomas S. Ulen, Assistant Professor,
Department of Economics

#599

Summary

The accepted theory of cartel behavior is reviewed both graphically and algebraically, a monopolistically competitive industry and firm serving as the model. The theory is then criticized on four points: inattention to the capacity of firms; casting a dynamic problem as if it were one of comparative statics; failure to deal with entry by new firms or investment rivalry among existing firms; and failure to deal with internal enforcement devices by a cartel. Each point, when incorporated into the standard theory, is shown to alter significantly the received conclusions about cartel behavior.



We can stand a good deal of cheating better than competition.

John Murray Forbes, President of the Chicago, Burlington & Quincy Railroad in the late 19th century.

Modern economic theory has in principle agreed with Adam Smith that "People of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices."¹ But modern theory has further directed its powerful tools to the task of elucidating what the results of such conspiracies are likely to be both for consumers and for the conspirators. In doing so, it has offered predictions which might have surprised Adam Smith, viz., that competition is inevitably a stronger force than collusion. The means of establishing this hypothesis are elegant, and there is also much of substance to admire in the theory. These attributes do not, however, prevent the received theory of collusion from being incomplete. That the hypothesis is misleading is the central claim of this chapter and, in a more thoroughgoing way, of this entire work.

I shall address myself in section I to an examination of the theory as it is currently expounded in most microeconomics textbooks. There has, of course, been more advanced work on collusion; these works are not dealt with explicitly here, although a few are mentioned. The reason for this neglect is that the advanced literature has not really questioned the tenets of the simpler theory. The received theory--which I shall

¹Adam Smith, The Wealth of Nations (New York: Modern Library, 1937), p. 130.

call the primitive theory of cartels--is treated both graphically and algebraically in the first section. Section II is a critique of the received, primitive theory laid out in the first section.

I.

There are, in the literature, two central propositions made regarding cartels. The first is that in an industry whose unfettered condition is competitive there is an incentive to collude in order to establish a collective monopoly. This incentive can be demonstrated for the case of a perfectly competitive industry which is initially in long-run equilibrium, as is the case at P_c , Q_c in figure 1. Each of the n firms in the industry is producing $q_c = Q_c/n$ at the point at which the market price is just equal to both the marginal and average costs of production. Thus, economic profit for each firm is zero. The corresponding decisions for the individual firm are shown in figure 2. If the firms in this competitive industry, now at a long-run equilibrium, were to get together and agree to restrict industry output to $Q_m = nq_m$, then they might all share in the resulting monopoly profits, represented by the shaded area in figure 1. For the individual firm, belonging to the collective monopoly involves adhering to the collectively-determined quota, q_m , which is, of course, less than its output in the competitive case, q_c , as shown in figure 2. At the industry level the restricted output has also resulted in the higher market price, P_m .

The second proposition about cartels in this primitive theory is that, once the collusion is established, there is an irresistible incentive for each member of the cartel to cheat, that is, to violate its quota. In figure 2 it is obvious that the firm is making greater

3.

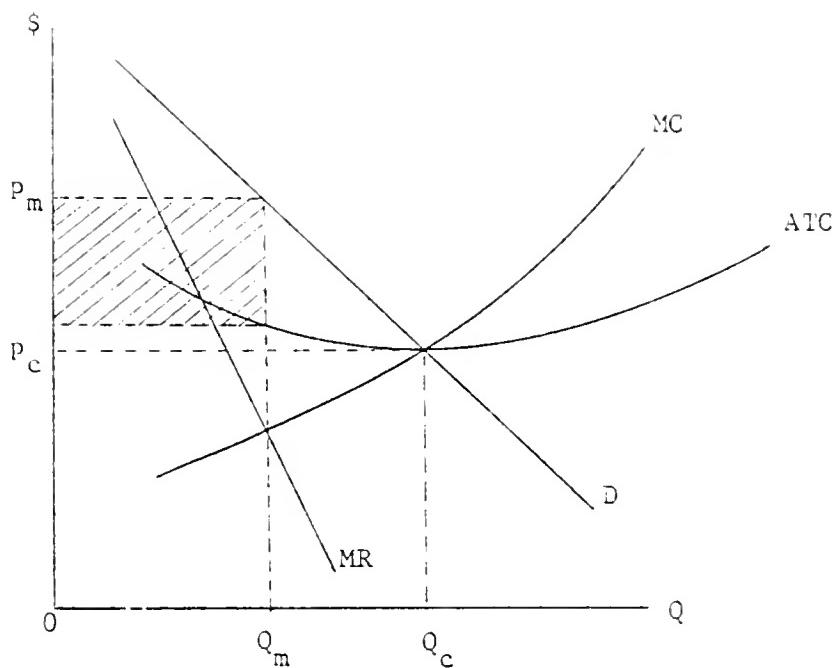


Figure 1

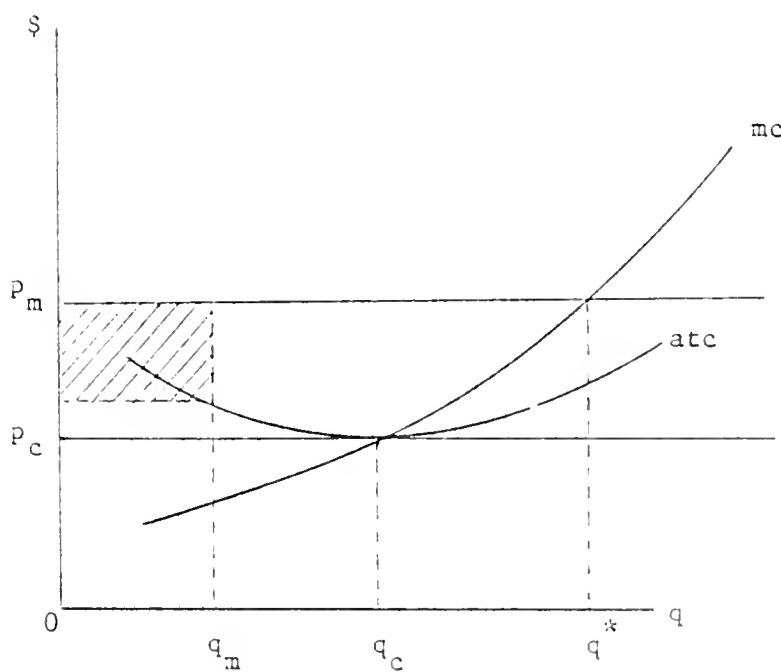
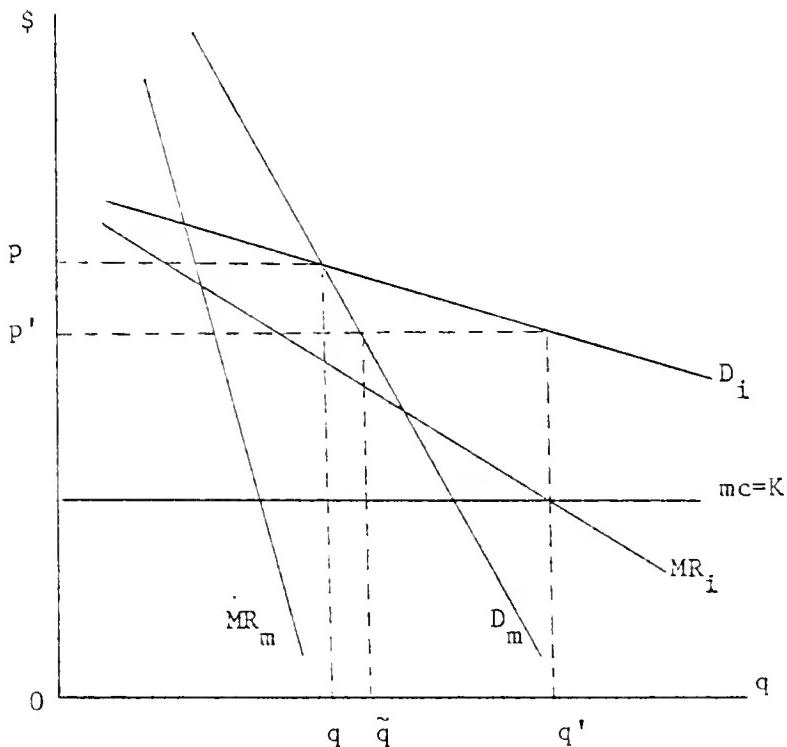


Figure 2

profits at P_m , q_m than it was at P_c , q_c . But it is also obvious that for the market price P_m the profit-maximizing output rate for the firm is not q_m but q^* . Since it is true that each firm perceives this incentive to cheat on its quota, each firm will surreptitiously produce q^* . The inevitable result is an excess market supply at the cartel price P_m equal to $n(q^* - q_m)$, which drives the market price down to its old competitive rate. That is, the competitive equilibrium will dominate any other in this industry.

These propositions can also be shown to hold for the slightly more complicated and more realistic case of an imperfectly competitive industry. Imagine that a firm's situation in such an industry is represented by the conditions in figure 3. The firm's marginal costs are assumed to be constant and equal to K .^{1a} As usual, the firm perceives two different demand curves facing it: one is simply its market share demand curve, D_m . Price and quantity pairs read off D_m are for the case in which all firms in the industry are quoting the same price. Associated with D_m is a marginal revenue curve, MR_m . The horizontal sum of the D_m 's for all firms in this imperfectly competitive industry would give the industry demand curve. In addition to D_m each firm also perceives the more elastic demand curve, D_f . This second curve tells the firm what the demand for its product will be when the firm alters its price while all other firms in the industry maintain theirs. The commonly accepted explanation for the greater

^{1a}Very similar results would result from assuming the usual upward-sloping firm supply or marginal cost curve. The constant case is chosen here for several reasons: it is easier to deal with without significantly altering the results; and it is consistent with the notion of economies of scale, which are thought by many to have been the short-run rule in late 19th century.



$q - \tilde{q}$ = the increase in sales occasioned by the secret price cut and due to increased purchases by old customers or first purchases by customers new to the industry,

$q' - \tilde{q}$ = the increase in sales attributable to customers stolen from other firms in the industry.

Figure 3

elasticity of D_i is that, when the firm secretly cuts its price, all others holding theirs at the higher level, it increases its sales from two sources. The lower price not only encourages old customers to increase their purchases or new customers to enter the market but also attracts to the cutting firm some customers who are switching from another firm in the industry to the cutter. As a very rough approximation, these two groups may be seen in figure 3. The firm cuts its price from P to P' , all other firms maintaining P , and increases its sales from q to q' . Note that this increase is clearly greater than what would have been the increase in sales had all firms cut their price from P to P' . In this latter case the firm's new sales would have been simply \tilde{q} . It is the presence of output heterogeneity and consequent brand loyalty among consumers which necessitates the theoretician's positing this second, more elastic, demand curve, D_i . A further point worth noting is that, while D_m is, in the short-run at least, assumed to be given and fixed, D_i is but one of a family of demand curves which pass through each point on D_m and tell the firm how much it can sell for any secret price cut whatever the price quoted by all the other firms in the industry.

In figure 4 the firm is in a long-run equilibrium, in which price is equal to average costs, the industry is charging P_c for its output; and the firm pictured is selling q_c , with no incentive either to charge a different price or to alter its rate of sales. Each firm is earning zero economic profit.

Now suppose that the firms in the industry undertake to form a cartel. Is there an incentive for the firm to join? By restricting

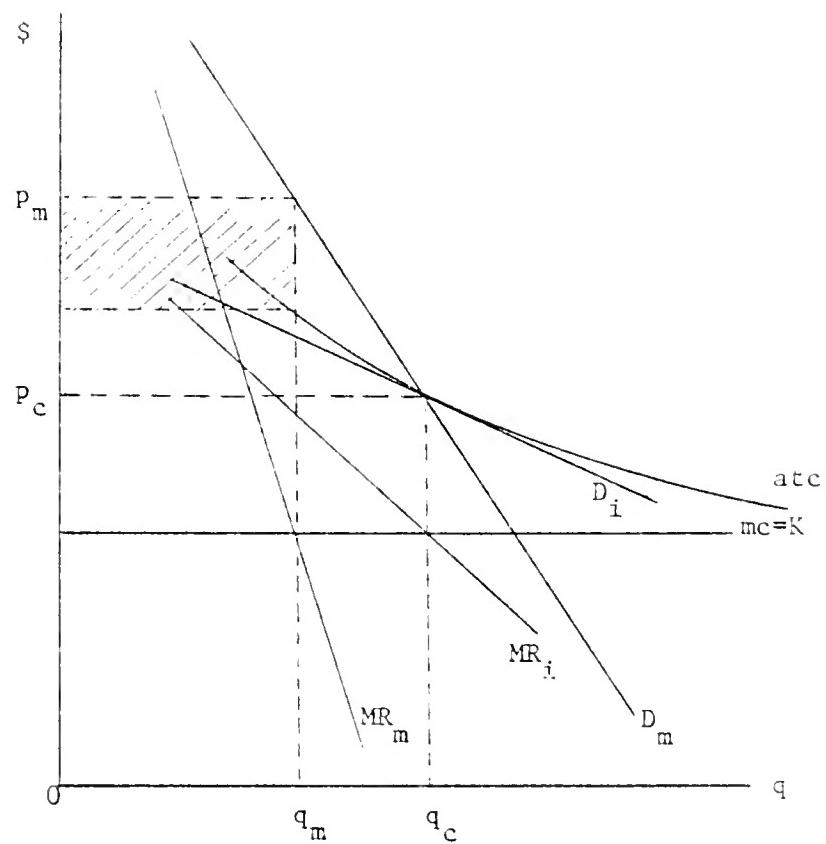


Figure 4

total industry output through the assignment of quotas to each firm, the industry-wide price may be raised. For the firm the situation is again shown in figure 4. By accepting the quota q_m , for which marginal cost equals marginal revenue from the market share demand curve, the output price rises to P_m with the result that economic profits in the amount of the shaded areas are earned by each firm. Therefore, there is an incentive to enter into a cartel or collective monopoly in an imperfectly competitive market.

In figure 5 the firm is shown at the price, P_m , and output, q_m , which have been determined to be the joint profit-maximizing ones for the cartel. The firm's average cost curve is labelled ATC, and its profits at the cartel price are indicated by the shaded rectangle labelled A. Can the firm do better than this under the circumstances? The answer is that the firm could increase its profits by disregarding both the cartel price and the firm's quota. The crucial assumption for this to be true is that the firm believes that it can successfully offer a secret price reduction which its co-conspirators will not immediately discover and attempt to match. In figure 5 it is clear that even the slightest shading of the cartel price will increase the firm's profits. Given that all the other firms are expected to continue to quote P_m , what is the optimal price for the cheater to offer? That optimum is P_1^* for which quantity $MR_1 = K$. The profits which the cheater earns at this combination (P_1^* , q_1^*) are labelled B and are clearly greater than the profits--the rectangle A--which accrued from loyalty to the cartel.² Thus, the second proposition about

²The possibility that B < A will be addressed later.

9.

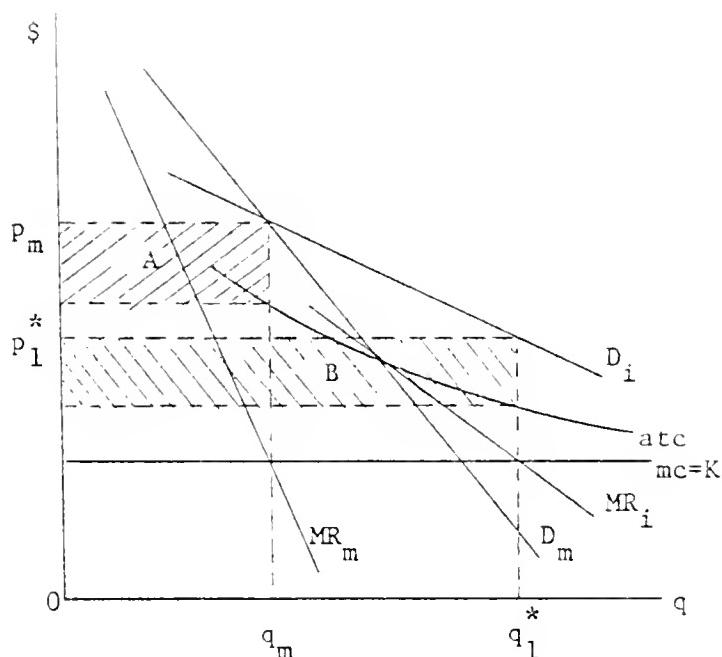


Figure 5

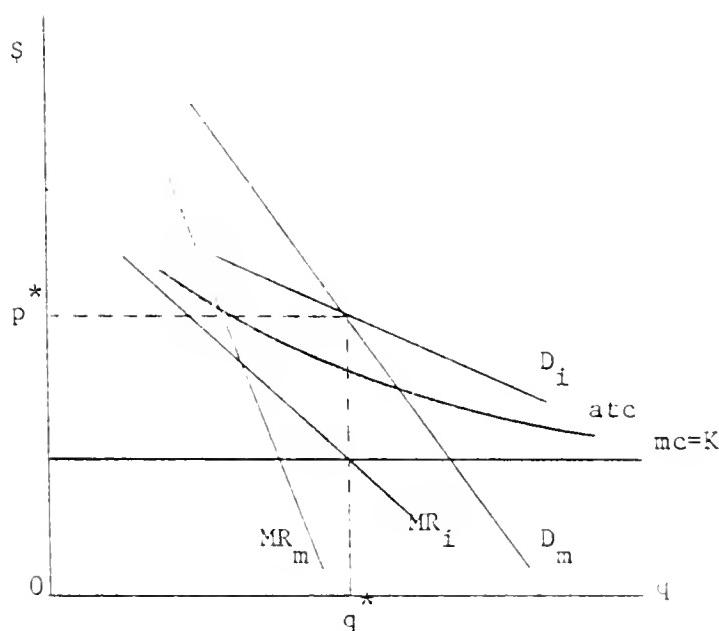


Figure 6

cartels--that there is an incentive to cheat once the collusion is established--is as true for an imperfectly competitive market as it is for the perfectly competitive case. It should also be evident that, because of the validity of the second proposition, the competitive equilibrium dominates in this market structure just as it did in the previous case. There are two different means of establishing this in the present case. One might assume initially that what is evident to the firm sketched in figures 4 and 5 is equally evident to all firms in this market. Therefore, they will all violate both the cartel price and their quotas. The result will be an excess market supply. The market price will drop until the firms have no further incentive to alter their prices, a situation which in the short-run looks, for the firm, like the one in figure 6 and in the long-run like that at (P_c, q_c) in figure 4. Alternatively, and perhaps more realistically, one might assume that the advisability of cheating is obvious at first only to one firm. Only after some time does the non-adherence of one member become obvious to the others, who, in the primitive theory, take action against the cheater by matching his discount. What happens next is shrouded in mystery. It seems just as likely that the truant will return to the fold as that he--or someone like him--will cut his price again and initiate a new round

of secret discount/matching, ending in the competitive equilibrium.³

I shall return to this possible ambiguity in the following section.

Let us assume that the market share demand curve, D_m , can be written as

$$(1) \quad p_m = \alpha - \beta q_m,$$

that is, as a linear demand curve with intercept α and constant slope β . It follows that the more elastic demand curve D_i , which gives the demand for the firm's output when it alone alters the price, all other firms holding their constant, may be represented by

$$(2) \quad p_i = \alpha - \beta q_m - \gamma(q_m - q_i).$$

with $\alpha > \beta$ and $\gamma > 0$, $\beta > 0$. And assume that marginal cost is constant and equal to K . Let us suppose that the firm, which knows all these facts about its costs and demand, enters into a collusion with the other firms in its imperfectly competitive industry so as to maximize the joint profits of the cartel members. What will the firm's share of those profits be? Let π_m represent that amount. We know that by definition

³The issues raised by this alternative assumption of an initial lone cheater are perhaps the most perplexing in all the theory of cartels. As will be seen below, two attempts have been made to extend the primitive theory from the basis of this assumption. George Stigler in his 'Theory of Oligopoly' has asked when a member of a cartel can justly infer that one of his co-conspirators is cheating. MacAvoy and Orr have attempted to define the optimal way for the loyal members to band against a cheater. Neither of these is entirely satisfactory. There appear to have been several instances in the history of collusion in which a cartel discovered a cheater, matched the cut for a time, and then returned all together to the monopoly price.

$$(3) \quad \pi_m = TR_m - TC_m$$

$$= p_m q_m - \int_0^{q_m} K dq_m$$

$$= p_m q_m - p_m K$$

Therefore, substituting (1) into (3), we get that

$$(4) \quad \pi_m = \alpha q_m - \beta q_m^2 - q_m K.$$

To find the profit-maximizing output, q_m^* , we know that we need to find that output for which $\frac{\partial \pi_m}{\partial q_m} = 0$ and for which the second-order condition is also satisfied. Performing these calculations on (4) gives the following:

$$(5) \quad q_m^* = \frac{\alpha - K}{2\beta}$$

for which the associated price from D_m is

$$(6) \quad p_m^* = \frac{\alpha + K}{2}$$

Thus, substituting (5) and (6) into (4) gives

$$(7) \quad \pi_m^* = \frac{(\alpha - K)^2}{4\beta}$$

We must now inquire as to whether the firm could do better than π_m^* by not adhering to its quota, q_m^* , and the cartel price, p_m^* . I shall define the profits from cheating as being π_i , as follows:

$$(8) \quad \begin{aligned} \pi_i &= TR_i - TC_i \\ &= p_i q_i - q_i K. \end{aligned}$$

From the formula for D_i , equation (2) above, we have that

$$(9) \quad p_i = \alpha - \beta q_m + \alpha q_m - \beta q_i \\ = \alpha - q_m(\beta - \alpha) - \gamma q_i.$$

We should very much like to have that particular D_i which passes through the point (p_m^*, q_m^*) on D_m since it is from that initial point that decisions about adherence versus non-adherence will be made. If we substitute from (5) into (9), we shall not only get the D_i we seek but also make p_i a function of q_i and of the parameters α , β , and γ . This substitution gives

$$(10) \quad p_i = \alpha - \frac{(\beta - \gamma)(\alpha - K)}{2\beta} - \gamma q_i$$

$$= \frac{\alpha\beta + \beta K + \alpha\gamma - \gamma K}{2\beta} - \gamma q_i$$

Let (11) $T = \frac{\alpha\beta + \beta K + \alpha\gamma - \gamma K}{2\beta}$ and now

substitute into (10) in order to get

$$(12) \quad p_i = T - \gamma q_i.$$

Since the slope of the more elastic demand curve is greater than that of the market share demand curve, i.e., since $\gamma > \beta$, the intercept along the horizontal axis must be greater for D_i than for D_m , i.e.,

$T < \alpha$.⁴ From the equation (12) we can derive the formula for the marginal revenue curve associated with D_i , namely

$$(13) MR_i = T - 2\gamma q_i.$$

The cheater can maximize his profits--assuming all other firms continue to quote p_m^* --by choosing that output, in violation of his quota, q_m^* , for which

$$(14) MR_i = MC = K$$

Setting (13) equal to marginal cost gives

$$(15) q_i^* = \frac{T - K}{2\gamma}$$

which is clearly analogous to (5), the formula for q_m^* . Furthermore, the optimal price to charge for the disloyal firm is

$$(16) p_i^* = \frac{T + K}{2},$$

⁴This is graphically obvious but can be shown through induction as follows:

$$(11) T = \frac{\alpha\beta + \beta K + \alpha\gamma - \gamma K}{2\beta}$$

$T > \alpha$ implies that

$$\frac{\alpha\beta + \beta K + \alpha\gamma - \gamma K}{2\beta} > \alpha.$$

$$\alpha\beta + \beta K + \alpha\gamma - \gamma K > 2\alpha\beta$$

$$\gamma(\alpha - K) > \beta(\alpha - K)$$

$$\gamma > \beta.$$

We have established that $T > \alpha$ implies that $\gamma > \beta$, which is what we have assumed to be the case for D_i to be more elastic than D_m . Since $T > \alpha$ implies something else which we know to be true, $T > \alpha$ must be true.

which is, of course, the counterpart to p_m^* as given in equation (6).

From (15) and (16) it is a simple matter to calculate π_i^* , the optimal profits for the cheater. This quantity is

$$(17) \quad \pi_i^* = \frac{(T - K)^2}{4\gamma}$$

Now, the question the cheater must ask himself is whether $\pi_i^* > \pi_m^*$.

Only if it is, is the cheating worth undertaking.

Recall that from (11) T was defined as follows:

$$(11) \quad T = \frac{\alpha\beta + \beta K + \alpha\gamma - \gamma K}{2\beta}$$

Therefore,

$$(18) \quad (T - K) = \frac{(\beta + \gamma)(\alpha - K)}{2\beta}$$

Thus, the optimal profits for the cheater may be written

$$(19) \quad \pi_i^* = \frac{(\beta + \gamma)^2(\alpha - K)^2}{16\beta^2\gamma}$$

This clearly facilitates comparison of cheating profits with those at the cartel optimum. If disloyalty is to be preferable to loyalty, then $\pi_i^* > \pi_m^*$, and this implies that

$$(20) \quad \frac{(\beta + \gamma)^2(\alpha - K)^2}{16\beta^2\gamma} > \frac{(\alpha - K)^2}{4\beta}$$

Upon simplification this inequality becomes

$$(21) (\beta - \gamma)^2 > 0.$$

It is clear that the only instance in which this condition will not be satisfied—that is, the only circumstances under which it will not pay to cheat rather than to remain loyal—is if $\beta = \gamma$. This can occur only if there is no monopolistic competition, that is, only if D_i and D_m are the same, and there is, therefore, no product heterogeneity. Given our assumption that $\gamma \neq \beta$, vis., $\gamma > \beta$, profits from cheating will be greater than those from adhering to the joint-profit-maximizing cartel price and its production quota.

Suppose that cheating occurs and that the cartel responds by quoting the same price as does the cheater. Could some member then quote a new, still lower, price and thereby earn greater profits? The answer is yes. Cuts will continue until the following conditions are simultaneously satisfied:

$$(22) p_i = p_m = p^*$$

$$(23) q_i = q_m = q^*$$

$$(24) MR_i = K.$$

It is only under these conditions that there is no incentive for any firm to charge a different price from that charged by any other firm. Clearly such a point would be the short-run equilibrium for any monopolistically competitive industry. How might we describe such an equilibrium? We have expressions for the conditions (22) - (24) so it is no trouble to derive the price and quantity configurations for

the competitive equilibrium. From the equilibrium conditions and

the expression for MR_1 , we have

$$(25) q^* = \frac{\alpha - K}{\beta + \gamma}$$

and

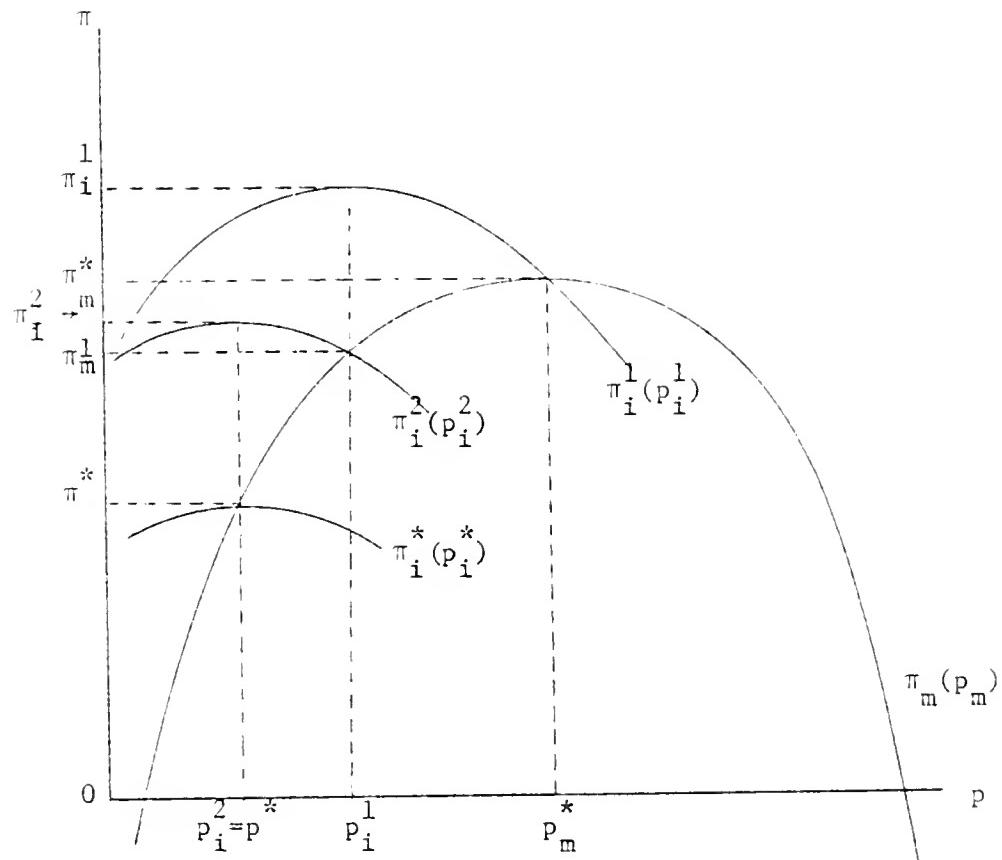
$$(26) p^* = \frac{\alpha\gamma + \beta K}{\beta + \gamma}^5$$

⁵The solution given here for the short-run equilibrium for an imperfectly competitive industry exhibits, as does no other treatment known to me, the often-expounded fact that the equilibrium price for this sort of industry will be somewhere between what it would have been had the industry been either monopolistic or perfectly competitive. We know that if this industry were perfectly competitive price would equal marginal cost, i.e., $P^* = K$. We have seen (equation (7) above) that if the industry were a monopoly, $P_m^* = \frac{\alpha + K}{2}$. The usual intermediate microeconomic contention is that

$$P_m^* > P^* > P_c^* \text{ or } \frac{\alpha + K}{2} > \frac{\alpha\gamma + \beta K}{\beta + \gamma} > K.$$

On the assumption that $\gamma > \beta$, it is an easy matter to verify these inequalities. It is a matter worth noting and worth bearing in mind in what follows that one of the rich lodes of monopolistic competition theory as yet unmined has to do with the relationship between D_1 and D_m , i.e., between γ and β , the slope parameters for those demand curves. A moment's reflection clearly suggests that γ and β are not independent. The degree to which one firm's output is distinguishable from that of another firm is among other things what determines the relationship between γ and β . And clearly this distinguishability is something over which the firms' themselves have some control. An example of this point is given by the now-booming running shoe industry. There are some firms who boldly distinguish their product from the familiar three stripes of Adidas. Nike, for instance, uses a large yellow swoosh on the side of their shoe. Other manufacturers have similarly distinctive markings on their shoes. However, there are some manufacturers e.g., Sears, K-Mart, and J.C. Penney, who choose to minimize the difference between their product and that of Adidas by putting two or four stripes on their product. Interestingly, the firms seeking distinction charge very similar prices to those of Adidas. Those trying to camouflage differences usually offer a lower price than does Adidas.

The relationship between the incentive to collude and the incentive to cheat may perhaps helpfully be viewed with the aid of the accompanying graph, figure 7. I have graphed profit on the vertical axis as a function of price, on the horizontal axis. The parabola marked $\pi_m(p_m)$ is the profit contour along the market share demand curve, D_m . Initially profits are negative for very low prices although they are even then an increasing function of price. Only when price has been raised to that point at which marginal revenue is equal to marginal cost--at the price p_m^* --are profits at a maximum. Beyond this point any attempt to boost the price still further results, of course, in a diminution of profit. Recall that, since $\pi_m(p_m)$ is the profit contour associated with the demand curve D_m , that contour represents those situations in which all the firms are charging the same price. There is, in addition, a series of parabolas mounting the left side of the $\pi_m(p_m)$ contour and labelled $\pi_i^j(p_i^j)$. This series of contours corresponds to the various demand curves D_i^j which pass through any given point along D_m . I have drawn in only three of these $\pi_i^j(p_i^j)$ for very specific reasons. $\pi_i'(p_i')$ passes through the peak of the $\pi_m(p_m)$ hill. It therefore corresponds to the more elastic demand curve which, in figure 5, I have drawn through the joint-profit maximizing point (p_m, q_m) . A cheater operating along his own D_i from that point is thus moving along the parabola $\pi_i'(p_i')$. The third of these contours, $\pi^*(p^*)$, reaches its peak exactly on the grand contour $\pi_m(p_m)$. Note that all of the other lesser parabolas reach a maximum to the left--i.e., at a lower price-- $\pi_m(p_m)$. The second contour, $\pi_i^2(p_i^2)$, is simply intermediate between the other two.



$\pi_m(p_m)$: the profit contour along the market share demand curve, D_m .

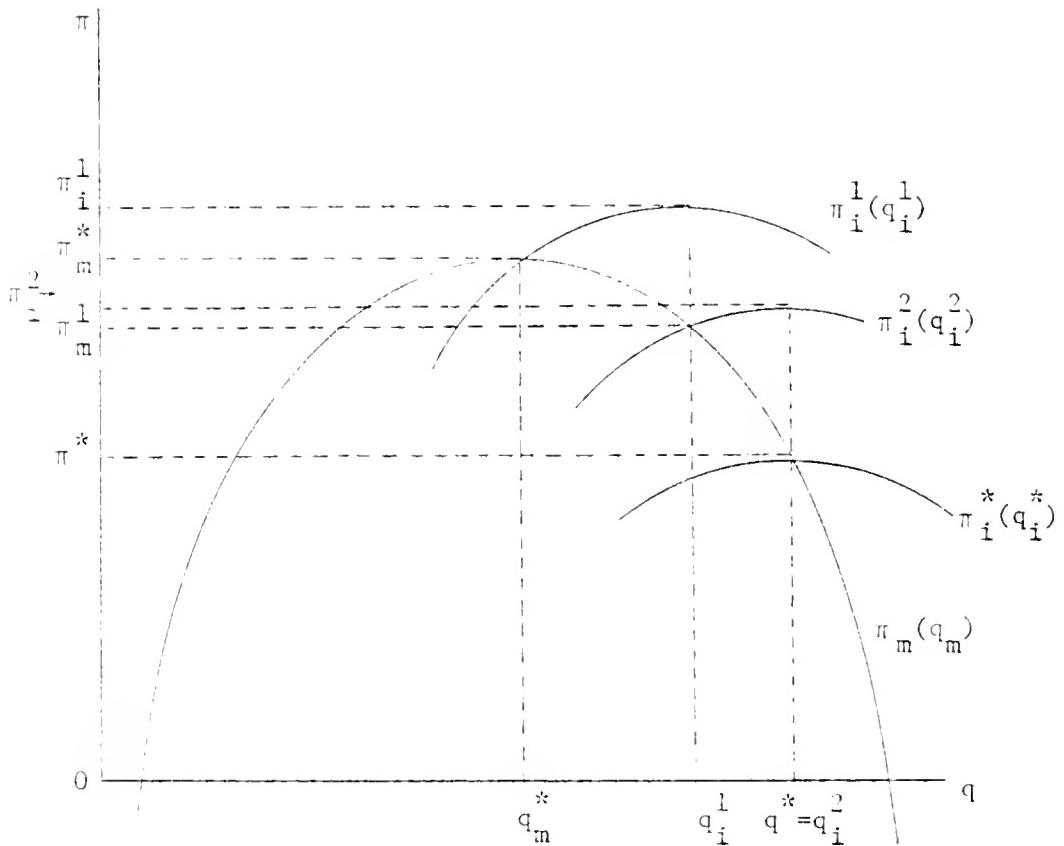
$\pi_i^j(p_i^j)$: the profit contour along the cheater's demand curve, D_i ,
at any given price along D_m .

Figure 7

We can now run through the life history of a cartel with the aid of figure 7. Suppose the cartel sets the joint profit-maximizing price, p_m^* , and therefore assures each firm in this monopolistically competitive industry profits in the amount π_m^* . The cheater perceives that he can do better than this. By moving along his own demand curve, D_i , quoting a lower price, and violating his quota, the treacherous firm can select the price p_i' which, given that all others are still charging p_m^* , will earn him greater profit, π_i' . When he is discovered, the cartel members--I assume--match his price. This shifts the cheater's demand curve D_i down along D_m to p_i^1 and places him back on the $\pi_m(p_m)$ at the point at which it intersects with $\pi_i^2(p_i^2)$. All firms are now charging p_i^1 , and each firm is earning profits in the amount π_m^1 . This is clearly less than what the firm earned when the cartel was established-- π_m^* --and when it first undertook to cheat-- π_i^1 . Let us now assume that from the position (π_m^1, p_i^1) it does not occur to the firms to return to the letter of the cartel agreement, i.e., to return to production of their quotas and to charging the joint profit-maximizing price. Rather they propose to remain at (π_m^1, p_i^1) for the time being. Assuming the initial cheater still desires another go at maximizing his profits, what will he now do? By cheating yet again he can earn slightly more than he could by remaining loyal. He therefore quotes a still lower price, p_i^2 , and earns π_i^2 , a greater sum than π_m^1 . When this round of deceit is uncovered, the cartel again matches the lower price. The cheater's demand curve is pushed down D_m to p_i^2 , which, I assume for simplicity's sake, is equal to p^* . The new cheating profit contour $\pi^*(p^*)$ reaches its maximum on $\pi_m(p_m)$ at this point. There is

no further incentive to cheat: by charging a price different from p^* --assuming all other firms hold fast to that rate--the cheater will receive less profits than π^* . Thus, the industry is at that point in short run equilibrium. A symmetrical sort of tale regarding the incentive to collude giving way to the incentive to cheat can be told in profit-quantity space rather than profit-price space. This is done in figure 8.

In these tellings, whether algebraic or graphic, there seems to be no way out of the conclusion that cartels are not stable. The incentive to cheat is stronger than the incentive to collude. The empirical predictions of this theory are fairly straightforward--firms in essentially competitive industries will try to form collective monopolies, but those collusions will be relatively short-lived. If one were to investigate such an industry, one would observe an initial rise in market price and a lowering in the production of each firm in the industry. Simultaneous with these changes the profits of each firm would increase, perhaps as reflected in the prices of the firms' shares quoted on a stock exchange. Shortly thereafter--the timing here is uncertain--some firm will have discovered that it can do better than by scrupulously remaining loyal. Thus, one would observe, if not the actual cheater, that the cartel price was being shaded and that the amount of output being offered for sale in the industry exceeded the sum of the individual firms' quotas. The cartel agreement would then be, like the Danish laws in Hamlet's time, "more honored in the breach than in the keeping," and prices would be observed to fall to something like the competitive level, output also increasing.



$\pi_m(q_m)$: the profit contour along the market share demand curve, D_m .

$\pi_i^j(q_i^j)$: the profit contour along the cheater's demand curve, D_i ,
at any given price along D_m .

Figure 3

II.

The received theory of cartels seems to be weak on four points. First, it pays no attention to problems of capacity among the member firms. Second, the model, although one of comparative statics, attempts to deal with an essentially dynamic problem: the movement from the joint-profit-maximizing price and quantity to the competitive static framework some crucial assumptions have been made but left unstated. When stated, they appear to me to be unsatisfactory. Third, the primitive theory of cartels ignores the possibility of entry into the market or of investment rivalry among the colluders, and thus skirts the longer term problems which may face attempts to maximize profits jointly. And lastly the theory grants to the economic agents in the model a selective rationality for which there is no basis either in theory or in fact. By the term "selective rationality" I mean rationality with regard to some but not all of the aspects relevant to a given problem. How this notion applies to the members of a cartel organized under the received theory will be made clear below.

The first criticism mentioned has to do with the productive capacity of the individual firms involved in the collective monopoly. In order for a cheater to accomodate an increase in demand for his services, he must have either excess capacity at hand or a relatively high elasticity of supply of capacity. Let us deal with the second possibility first, using a railroad involved in a cartel as an example. Clearly, in the short-run, a firm which wishes to transport more goods than allotted him by the cartel must have spare rolling stock at hand. Suppose it does not. Then it can only accomodate more customers by

other means. It might, for example, purchase more cars from its suppliers or, alternatively, it might hire some rolling stock from some other railroad which has spares on hand. The firm might also think of transferring some of its own rolling stock from some other, less profitable routes to the route upon which it desires to cheat. Lastly, if possible, it might attempt to load its existing cars more densely at the cost of this increased weights' heightening wear on the rails and boosting the fuel requirements of the locomotives. These alternatives are all costly to one degree or another. An important point beyond that is that the visibility to the firm's rivals of these various means of increasing capacity differs. Purchase or rental is very likely to be evident to the firm's co-conspirators and thus to put them on their guard. The other capacity-incrementing schemes are less visible to rivals but also fraught with other problems, like precise calculations of profit rates on competing routes. And against these differing supply costs there is the likelihood of the cheating's not succeeding.

The point to which these examples are directed is that cheating on a collusive arrangement is not costless, as the received theory implicitly assumes. Rather, the decision to violate one's quota is a decision to incur certain costs, some of which appear to be related to the desire to accommodate the extra-quota business through an increase in capacity and others of which appear to be related to hiding this decision from the cheater's co-conspirators.

The suggestion here is that the primitive theory of cartels implicitly assumes that the supply of capacity to the firms in the cartel

is perfectly elastic. This seems as if it might be a restrictive assumption and urges one to investigate the effect on the received model of assuming the contrary, namely, a capacity constraint being imposed on the members of the collusion. This possibility is shown in figure 9. There the capacity K of the firm corresponds to output \bar{q} . At that level of output the marginal cost curve displays a discontinuity by becoming exactly vertical.⁶ The figure is drawn so that the marginal revenue curve associated with the firm's market share demand curve passes through the discontinuity in the marginal cost curve. This would be the case if that cartel were fortunate enough--given the level of industry demand and each firm's capacity--to assign as a quota to one firm exactly the upper limit of what the firm is capable of producing. If all the members are in the same position, that is, for all firms $\bar{q} = q_m^*$, the cartel will be absolutely stable in the short-run. The desire to cheat may be there, but, in the short-run, the wherewithal to do so is missing. Thus, it would appear to be the case that the less elastic the supply the more likely the cartel is to be stable.

Let us explore this possibility further by considering the case, shown in figure 10, in which $\bar{q} > q_m^*$; the firm has excess capacity in the amount ($\bar{q} - q_m^*$). Clearly the firm has the usual incentive to cheat from its initial output rate of q_m^* . I assume that the firm does so by selecting that output for which MR_i' is equal to MC . As

⁶ A more realistic depiction would have the marginal cost curve begin to rise as it approached \bar{q} and thereafter become steeper and steeper. The refinement which this would allow does not warrant the complications which it introduces.

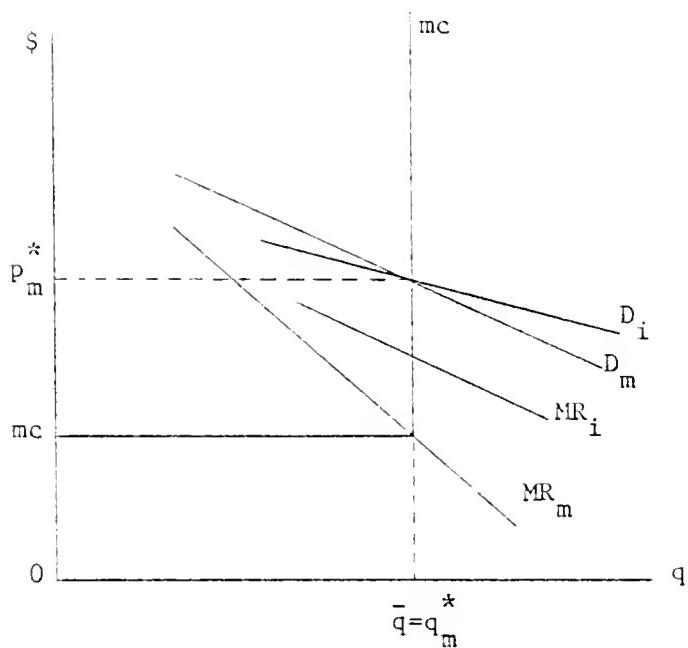


Figure 9

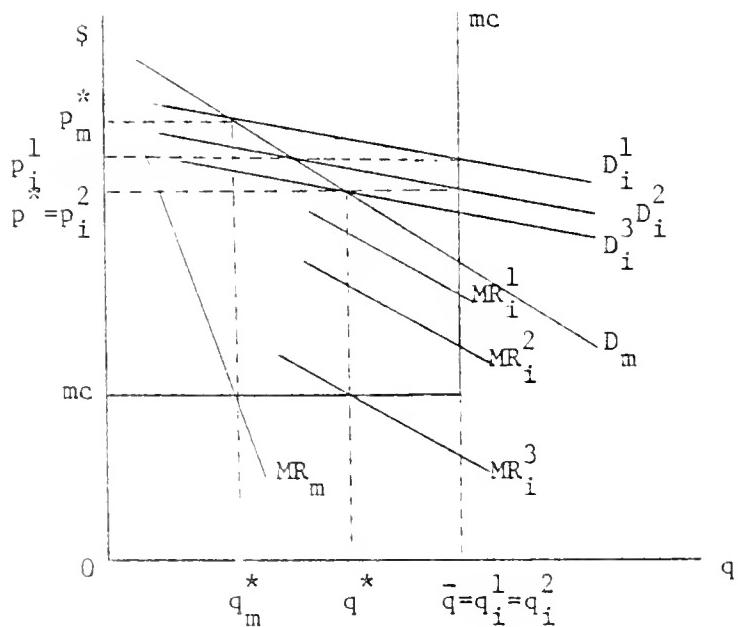


Figure 10

the discounted price of p_i^1 . Note that if its capacity were larger, this firm would have chosen to offer a greater amount of output at an even lower discount. As before, I assume that the cheating is discovered and that the price cut is matched by the other members. This leads to a downward shift in D_i to D_i^2 . A new round of cuts ensues leading to the short-run industry equilibrium along D_i^3 .

This second example confirms the prediction above that the virulence of the cheating will be somewhat mitigated by the presence of an upper bound on firm output. This is specifically true only when $\bar{q} < q^*$, where, as before, q^* is the short-run monopolistically competitive output and \bar{q} is the short-run capacity of the firm. Obviously the tempering effect of this output constraint on cartel well-being is greater the less is the difference between a particular member's quota and his capacity.

This point suggest two others. First, if it is at all possible, a cartel can increase its chances for success (in the short-run at least) by so assigning quotas as to minimize excess capacity for each member. Secondly, any event which causes a reduction (increase) in the excess capacity of the members of a collective monopoly will increase (decrease) the organization's chances for super-normal profits. Clearly, one such event would be fluctuations in the demand for the industry's output.

The second criticism which should be raised against the received theory has to do with the fact that there is something too mechanical about the fact that departures from the joint-profit-maximizing price and quantity inevitably lead to the short-run competitive equilibrium.

Imagine the following happening: from the initial cartel price and quota a member perceives the incentive to cheat and offers a secret discount. In time he is found out, and all the other firms match the lower price. We have seen that from this point someone now perceives that rather than remaining at this price-quantity combination, it again pays to cheat in that still greater profits can be earned by being disloyal. But these greater profits from a new round of cheating depend on the cheater's believing that a new price cut will not be perceived for a time and, thus, not be matched immediately by all other firms. Whereas these might have been plausible beliefs just after the formation of the cartel when all were flush with the newness of the thing, these now seem, in view of the recent experience of disloyalty and detection, naive beliefs. Everyone will be on the lookout for cheating. It is thus not unlikely to be quickly found out and just as quickly matched. This fact surely reduces the incentive to embark on a new round of secret cuts. But the crucial point is, how do these shrunken returns from a second round of cuts compare to the profits from returning to the original joint-profit-maximizing situation? Might it not be a wiser thing to opt for the surer cartel profits? But then these have been somewhat reduced in value by the discovery that cheating will occur. Still, without a return to the cartel situation the end of the line will be the competitive equilibrium, which, while stable, returns the least profits imaginable. It might be a wiser course to try--after the first round of cheating is discovered and answered--to re-establish the cartel, even though one might expect periodic cheating in the future.

John Murray Forbes, the Coston capitalist and investor in the Chicago,

Burlington & Quincy Railroad, put the matter as follows: "We can stand a good deal of cheating better than competition." I am not suggesting that an episode of disloyalty will always lead to a resurrection of the collective monopoly optimum; merely that it might and that this possibility is suggested by the same sort of reasoning which in the received theory hastily plunges the cartel into a stable, competitive equilibrium.

The third shortcoming from which the primitive theory suffers is its failure to deal with the possibility of entry into the market in the long-run or of investment rivalry, also in the long-run, among the members of the cartel.⁷ Surely the general tendency to ignore these problems is attributable to the conclusion of the received theory that the cartel agreement can be expected to be broken in the short-run. If, during that time period, the industry can be shown to return inevitably to the competitive equilibrium, then the long-run equilibrium will simply be that of a perfectly or imperfectly competitive industry, as the case may be. On the other hand, if we grant the possibility that the cartel may be successful in restricting total industry output and in maintaining the monopoly price in the short-run, then it becomes evident that there may well be a distinction between the analysis of the cartel in the different time periods.

Assume that the cartel has discovered the proper means to minimize the incentive to cheat in the short-run or that, alternatively, each firm in the collusion is operating at its capacity constraint so

⁷ The only treatment of this latter subject with which I am familiar is that in George Stigler, The Theory of Price, 3rd ed. (New York: The Macmillan Co., 1966), pp. 234-36.

that no firm has the wherewithal to cheat. The monopoly profits earned by the essentially competitive industry under these circumstances will certainly encourage resources into the industry or the existing firms to expand their capacity. Thus, the problem for a successful cartel is how to prevent entry or investment rivalry in the long-run. These are different matters, which is not to say that they may not happen simultaneously. It may well be, that is, that entrants may appear at the same time that the members of the cartel are attempting to put additional capacity in place. Inasmuch as these are different sorts of threats to stability in the collusion, they require different solutions, and, for the economist interested in modeling the collusive process, different analytical treatments.

There is not a commonly agreed-upon method of precluding firms from competing among themselves in the size of their fixed capacity. Some schemes for doing so are, however, conceivable. For instance, the members of a cartel might designate spheres of influence for themselves and agree not to intrude into each other's territory. The hope would be that each firm would confine its investments to its own territory and, thus, not threaten the stability of the collusion in the commonly-served territory. This method of defusing long-run competition is more likely to work the larger the total area served by the cartel members in relation to the area over which the collusion operates.⁸

⁸Granting territorial exclusivity was a common means of promoting stability in late nineteenth century railroad cartels, as it is in modern industries, e.g., the hearing-aid market.

If the distinction between firms is not territorial but has to do instead with the range of products manufactured, then the agreed-upon restriction on investment might be to refrain from investing in plants designed to produce the cartelized output common to all the members. The hope would be that the members would invest instead in plant for the production of its non-cartelized output. If the capital or other fixed input happens to be used jointly to produce all of the firm's production, then this sort of attempt to restrict investment rivalry is not likely to be very effective.

The members of some collusions might be willing to pool their investment funds and relinquish control over investment decisions to a committee appointed by all the members of the collusion.⁹ While this seems the ideal solution to the problem, it demands an unlikely commitment from the colluders, viz., the sacrifice of firm autonomy.

Even if none of these solutions is possible, it must be recognized by all colluders that the commitment to expansion of one's plant is a highly visible means of signallying one's intention with respect to the other members of the cartel. For that reason it would seem likely to posit that investment rivalry, precisely because it is such a strong threat to the stability of the cartel, will be undertaken only when other events have indicated that the collusion is weakening. That is, just as war can be taken as the signal that all other means

⁹ Some fixed assets of the Southern Railway and Steamship Association were jointly owned by the members of the cartel. See Thomas S. Ulen, Cartels and Regulation: Late Nineteenth Century Railroad Collusion and the Creation of the Interstate Commerce Commission (Unpublished Ph.D. dissertation, Stanford, 1978), Appendix to Ch. 2.

of resolving a conflict have proved unsuccessful, so investment rivalry within a cartel is probably a signal that something else has gone wrong and that the members have chosen to give up on making the alliance work. The power of this signal may, therefore, make it unnecessary for there to be an explicit agreement among the members regarding their plans to invest in fixed plant. All firms may take this as an instrumental variable for the trust which must govern their relations if the collusion is to work.

The cartel must also devise a strategy for discouraging entry into its industry. This is so because the more successful the cartel is in the short-run in jointly maximizing profits, the more likely it is to attract resources to the production of the cartelized product. Entry may be extremely costly because of the size of fixed plant involved, in combination with essential imperfections in the capital market. If, however, there are no such constraints on a potential investor, the commonly used scheme for discouraging entry is to practice limit-pricing. The monopoly or collective monopoly does not set its price so as to maximize short-run profits, there being no surer way to attract entrants. Instead, the cartel charges a slightly lower price, trading off short-run gain for a longer term of market power in the industry, free from the competition offered by entrants. It has been argued, for example, that OPEC is not charging a price for oil which would truly maximize that cartel's short-run profits. The reason is that such a high price would heighten the incentive of oil consumers to adopt alternative fuel sources or of entrepreneurs to search for as yet unknown oil deposits or of innovators

to undertake research in currently unpractical energy technologies such as nuclear or solar power. To forestall these developments the members of OPEC can be said to be practicing limit-pricing, charging a price above the marginal and average cost of production and yet below that which a monopolist would charge if he were unafraid of entry.

It has been suggested that there is no need for a collusion or a monopolist to practice limit-pricing. This is because it can deter entrants through the means of keeping on hand enough excess capacity to be able to lower the output price very rapidly through expanding output. The potential entrant, who may have begun to lay plans for bringing resources to the industry in the belief that the monopoly price would be maintained, may suddenly find, after the expansion of output by the cartel or monopoly, that the price is too low to justify his going ahead with his plans. This lowering of the monopoly price need not, of course, be practiced time and again as new entrants threaten. It would probably be the case that the sudden lowering of output price need be demonstrated only once or twice for its effect to become evident to all future entrants.¹⁰

Assume that in the cartelized industry capacity is an upper bound on output and profits are given by

¹⁰An appropriate analogy to the value of excess capacity in deterring entry for a cartel might well be taken from the writing on slavery. Clearly the whip need only be applied in a vicious way once or twice for its fear to rise up in the slaves' minds and discourage them ever after from misbehaving. Thus, one might infer that the less the whip is used, the more effective it has been in controlling behavior.

$$\pi(x, k) = R(x) - c(x) - rk,$$

where x = output

k = capacity

$R(x)$ = the revenue function

$c(x)$ = the variable cost function, and

r = the rate of interest.¹¹

Average total costs are then given by

$$ATC \equiv a(x, k) = \frac{c(x)}{x} + \frac{rk}{x}$$

If $k = x$, then the firm is efficient in that costs are minimized for a given output. In that instance,

$$a(x, k) = \frac{c(x)}{x} + r$$

Suppose now that an entrant appears and supplies output y . This implies that total industry output is $(k + y)$. The existing firms have expanded output to $x = k$ and lowered their price to $P(k)$. (See the accompanying figure.) With the entrant's output the industry price falls further, to $P(k + y)$. If $a(y, y)$ is the average total cost curve for the entrant, then entry is deterred if for all y ,

$$\pi_y < 0.$$

That is, for all y , if

¹¹The notation and the accompanying graph are from Michael Spence, "Entry, Capacity, Investment, and Oligopolistic Pricing," Technical Report No. 131, (April, 1974), Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, Calif.

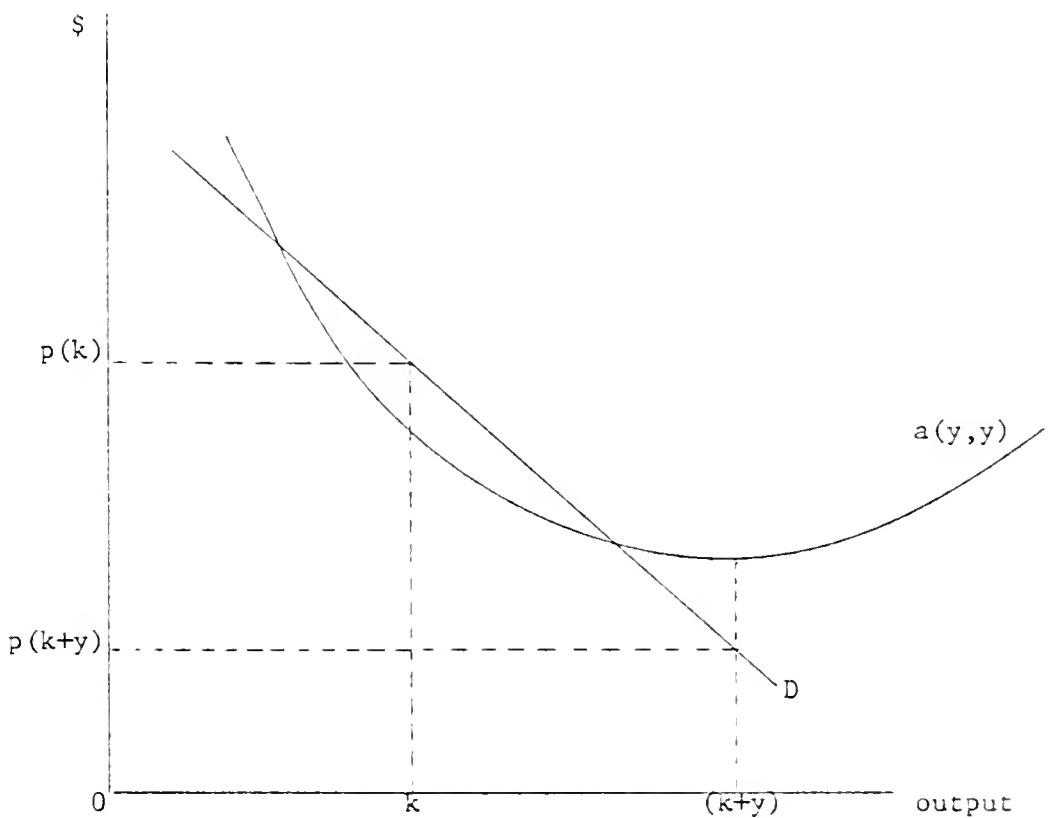


Figure 11

$$P(k + y) \leq a(y, y) = \frac{c(y)}{y} + r$$

as k increases, $P(k, y)$ falls for each y . For k sufficiently large, the residual demand is zero. If we set \bar{k} equal to the minimum capacity for which entry is deterred, then the industry maximizes its profits subject to the following two constraints:

- (1) $x \leq k$, which implies output is no larger than capacity, and
- (2) $k \geq \bar{k}$, which insures that entry is deterred.

The fourth criticism of the received theory regards what I consider to be the selectively-granted rationality in the model. The firms are given the wit to see that it pays to collude and that it pays to cheat, but they are not allowed to have the sense to put these two bits together in order to realize that their real problem is the prevention of cheating. The fact that cartels break down is (and has long been) no less evident to the members of a cartel than it is (and has been) to economic theorists.

Before the 1890's, when he became pre-eminent in the steel industry, Andrew Carnegie participated in price-fixing and market-share allocating arrangements with other firms in the industry. His lieutenant Charles Schwab remembered of these collusions that "...many of them lasted a day, some of them lasted until the gentlemen could go to the telephone from the room in which they were made..."¹²

¹² Jonathan R. T. Hughes, The Vital Few (New York: Oxford University Press, 1973), pp. 239.

Firms involved in cartels have therefore devoted resources to the task of reducing the incentive to cheat, once a collusion has been set up. The well-studied examples of the NCAA and the AMA both reveal that those important collusions are alive to the incentive to cheat and have concocted devices for reducing that incentive among their members. The NCAA employs private detectives to check on the offers which member colleges are tendering to prospective athletes. The series of public letters of intent and of acceptance which a recruit to an NCAA school must sign insure that the terms under which an offer has been made are nominally available for inspection by the officials of the cartel as well as by the other members. If, despite these efforts, a school violates the implicit wage constraint imposed by the collusion, and is discovered to be doing so, the cartel inflicts punishment--denial of access to lucrative television and radio contracts and to extremely profitably post-season games--which is severe enough to lesson the incentive to cheat. A similar set of policing and enforcement devices has been shown to be effective in dissuading doctors from violating their price-fixing agreement.





UNIVERSITY OF ILLINOIS-URBANA



3 0112 060296347